nature genetics

Gene expression–based high-throughput screening (GE-HTS) and application to leukemia differentiation

Kimberly Stegmaier[1], Kenneth N Ross[2], Sierra A Colavito[1], Shawn O'Malley[3], Brent R Stockwell[2] & Todd R Golub[1,2,4]

Chemical genomics involves generating large collections of small molecules and using them to modulate cellular states. Despite recent progress in the systematic synthesis of structurally diverse compounds, their use in screens of cellular circuitry is still an *ad hoc* process[1–4]. Here, we outline a general, efficient approach called gene expression–based high-throughput screening (GE-HTS) in which a gene expression signature is used as a surrogate for cellular states, and we describe its application in a particular setting: the identification of compounds that induce the differentiation of acute myeloid leukemia cells. In screening 1,739 compounds, we identified 8 that reliably induced the differentiation signature and, furthermore, yielded functional evidence of bona fide differentiation. The results indicate that GE-HTS may be a powerful, general approach for chemical screening.

To prove the feasibility of GE-HTS, we applied the method to the identification of compounds capable of inducing terminal differentiation of acute myelogenous leukemia (AML) cells. The plausibility of differentiation induction in leukemia is suggested by a rare subtype of AML known as acute promyelocytic leukemia (APL) in which treatment with all-*trans* retinoic acid (ATRA) results in clinical remissions through modulation of a mutated retinoic acid receptor alpha[5,6]. Unfortunately, ATRA has no clinical efficacy in other subtypes of AML[7]. But the fact that all forms of AML show a block of differentiation suggests that differentiation therapy might be possible, provided that the right compounds could be identified. The mechanism underlying such blocked differentiation is unknown in most cases; thus, it is not possible to carry out small-molecule screens against a validated target protein. Instead, a cell-based screening approach is needed. The usual approach to assaying the myeloid differentiation phenotype, however, involves a combination of visual inspection of nuclear morphology and biochemical tests, such as nitroblue tetrazolium (NBT) reduction[8], neither of which readily lend themselves to a high-throughput screening platform. We therefore sought to use GE-HTS to identify compounds capable of inducing the differentiation program in AML without first knowing the crucial targets of this process.

The first step in the GE-HTS procedure is defining the gene expression signatures of the biological states of interest. To accomplish this, we carried out oligonucleotide microarray-based gene expression profiling of pretreatment bone marrow samples derived from individuals affected with AML, as well as of fully differentiated peripheral blood neutrophils and monocytes derived from unaffected donors (gene expression data are available in **Supplementary Table 1a**,**b** online). We thereby identified genes correlated with the neutrophil-versus-AML or monocyte-versus-AML distinctions (**Fig. 1**). We selected from these differentiation-correlated genes a handful of marker genes that represented the diversity of myeloid differentiation yet could be easily measured in high-throughput procedures. Specifically, we selected the autosomal chronic granulomatous disease-associated gene *NCF1* and the gene encoding orosomucoid-1 (*ORM1*) as markers of the neutrophil lineage, and the genes encoding interleukin 1 receptor antagonist (*IL1RN*) and secreted phosphoprotein 1 (*SPP1*) as markers of the monocyte lineage. Glyceraldehyde 3-phosphate dehydrogenase (*GAPD*) served as an internal control because it showed minimal variation in expression across the entire microarray data set. We found that these five genes were appropriately regulated in the HL-60 cell line model of AML, in which cells differentiate into neutrophilic or monocytic cells after stimulation by ATRA or phorbol 12-myristate 13-acetate (PMA), respectively (**Fig. 1** and **Supplementary Table 1c**,**d** online). We therefore used the five-gene pattern as the differentiation signature for screening. These genes were selected not because of any causal role in the orchestration of the myeloid differentiation process, but rather because of their collective function as robust indicators of the differentiated state.

We next developed a detection assay for the differentiation signature that would be suitable for high-throughput screening. We carried out multiplexed RT-PCR of the signature genes from cells grown in 384-well culture plates. The PCR amplicons underwent a single base extension (SBE) reaction, and the five-plex SBE products were then subjected to matrix-assisted laser desorption/ionization time-of–flight (MALDI-TOF) mass spectrometry and showed quantitative detection across a working range of concentrations (correlation coefficient 0.92; see also **Supplementary Note** online). The GE-HTS

[1]Department of Pediatric Oncology, Dana-Farber Cancer Institute and Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]The Eli and Edythe L. Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02139, USA.[3]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02139, USA. [4]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA. Correspondence should be addressed to T.R.G. (golub@broad.mit.edu).

Published online 8 February 2004; doi:10.1038/ng1305

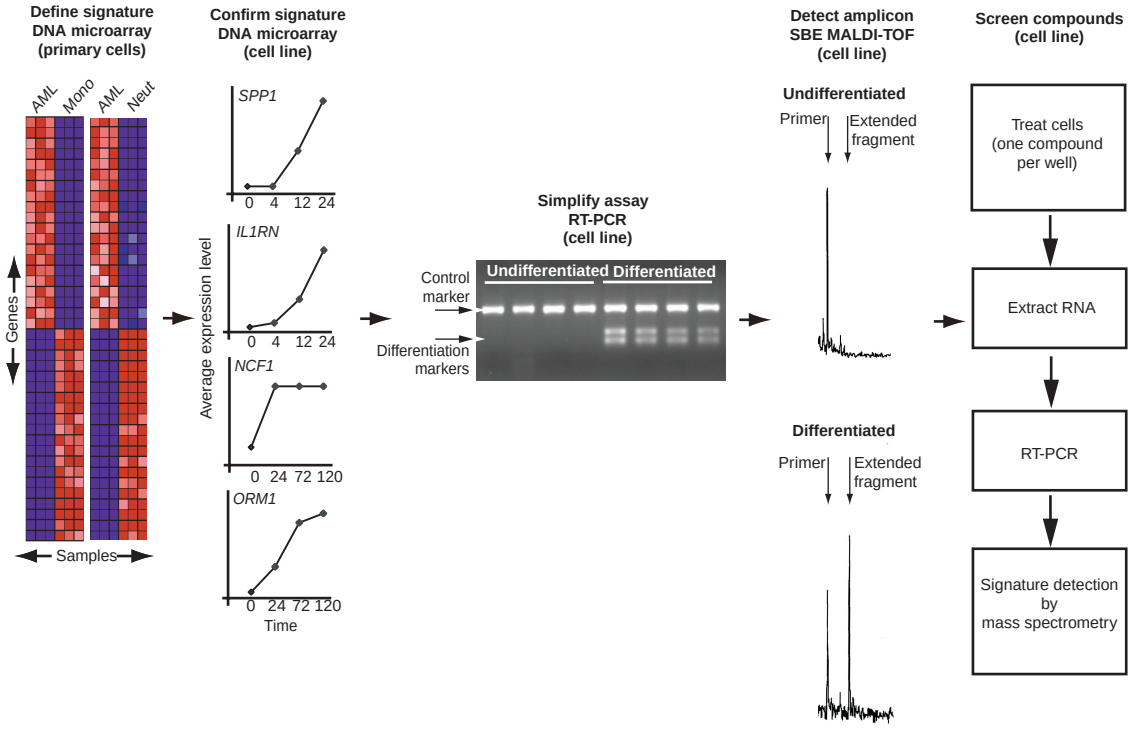

Figure 1 Screening schema overview. The gene expression signatures of the AML-versus-monocyte (Mono) and AML-versus-neutrophil (Neut) distinctions were first determined by DNA microarrays, as illustrated at left. Columns represent samples and rows represent genes, sorted according to their correlation with the distinction (red indicates high relative expression, blue low expression). Signature genes (*SPP1, IL1RN, NCF1, ORM1*) were then confirmed in an HL-60 cell line model of differentiation after stimulation by ATRA or PMA, and their differential expression (including *GAPD* as a control marker) extended to a multiplexed RT-PCR assay (middle panel). RT-PCR amplicons were then quantified by MALDI-TOF mass spectrometry after an SBE reaction. Shown is a portion of the mass spectrum of an undifferentiated HL-60 sample versus a differentiated sample corresponding to the *IL1RN* marker. In the undifferentiated sample, the predominant peak corresponds to the unextended SBE primer, whereas after differentiation (and expression of *IL1RN*), a peak corresponding to the extended fragment is observed. In the final step (right panel), cells were treated with one compound per well, their RNA was extracted and RT-PCR, SBE and mass spectrometry detection were carried out in 384-well format. The expression of the signature genes was thus quantified and used to generate a composite differentiation score from which hits were identified.

concept is not limited to the five signature genes used here or to the SBE–mass spectrometry detection method. Other high-throughput detection methods, such as bead-based detection or 'arrays of arrays' are, in principle, feasible[9,10].

For the small-molecule screen, we exposed HL-60 cells (18,000 cells per well in 384-well format) to a single compound per well in triplicate for 72 h. The compound library consisted of 1,739 chemicals that either were already approved for use in humans by the FDA or had already been extensively biologically characterized. We used 181 wells containing vehicle only as negative controls (complete library contents can be found in **Supplementary Table 1e** online). We carried out multiplexed RT-PCR for the five-gene differentiation signature and quantified the PCR amplicons by mass spectrometry as described above.

To score the results of the screen, we developed three tests of the signature genes. First, we analyzed the signatures with respect to *GAPD*

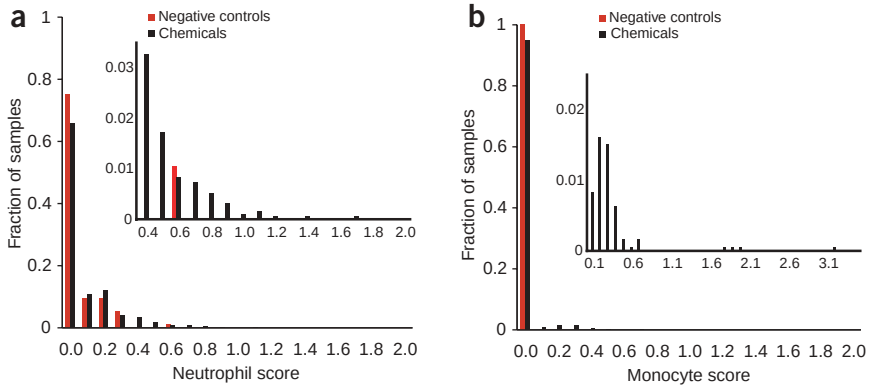

Figure 2 Distribution of differentiation scores in chemical screen. (**a**) Distribution of Neutrophil Scores resulting from triplicate screen of chemical library (see **Supplementary Table 1f** and **Supplementary Note** online for calculation of score). Red indicates distribution for 96 negative controls, and black indicates the 1,920 screened chemicals representing 1,739 distinct compounds. The inset shows the tail of the distribution in greater detail. A significant number of compounds induced high Neutrophil Scores, whereas the controls did not. (**b**) Data as in **a** for the composite Monocyte Scores.

Table 1 Summary of biological effects of candidate compounds

Compound	Name	Whole-genome *P* value	IC50 (μM)	NBT assay *P* value	Phagocytosis *P* value
A	(R)-(-)-apomorphine HCl	0.018	3.8	<0.001	0.034
B	4,5-dianilinophthalimide	<0.01	5.5	<0.001	0.005
C	Erythro-9-(2-hydroxy-3-nonyl)adenine HCl	0.017	>100	<0.001	0.205
D	5-fluorouridine	0.048	0.03	0.004	<0.001
E	16-ketoestradiol	0.026	>100	<0.001	0.455
F	Pergolide methanesulfonate	0.035	63	0.002	0.287
G	Cyclazosin HCl	0.662	7.3	0.004	0.010
H	1,10-phenanthroline	0.084	2.4	0.12	<0.001
I	All *trans* retinoic acid	0.013	0.4	<0.001	0.02
J	1,25-dihydroxy vitamin D3	0.011	9.1	<0.001	<0.001
K	Phorbol-12-myristate-13-acetate	0.025	<0.01	<0.001*	<0.001

The whole genome *P* value refers to genome-wide patterns of myeloid gene expression as estimated with the Mantel test. Using the Mantel test, we compared gene expression differences characterizing the original AML-versus-neutrophil distinction in primary cells with that of untreated versus chemical-treated HL-60 cells. We used a propidium iodine assay to determine the effects of compounds on cell proliferation and to calculate compound inhibitory concentration 50% (IC50). We used NBT reduction and phagocytosis assays to evaluate for biochemical and functional evidence of differentiation induction by chemical hits. Compounds I, J and K served as positive controls. Compound K reduced NBT reduction (asterisk) to less than that of undifferentiated controls.

RNA levels to exclude compounds that caused nonspecific cell death. Six percent of wells fell below a predefined *GAPD* score, consistent with previous observations that ~5–7% of compounds in this library result in cellular toxicity in multiple cell lines[11]. We also devised a Neutrophil Score and a Monocyte Score that incorporated the expression levels of all five signature genes and compared the distributions of these scores for the test compounds to those of negative controls. There was statistically significant skewing of the distributions of Neutrophil Scores ($P < 0.03$; **Fig. 2a**) and Monocyte Scores ($P < 0.04$; **Fig. 2b**) among the compounds tested.

We therefore examined 15 top-scoring compounds that emerged from the screen. Two of these candidates (aminopterin and dimaprit dihydrochloride) were previously reported as myeloid differentiation agents and were therefore not pursued further[12,13]. Repeat experiments with the 13 remaining candidate compounds indicated that 8 of them (**Table 1**) reproducibly triggered the differentiation signature (**Supplementary Table 1g** online).

We next asked whether these eight compounds simply regulate this handful of marker genes, or whether they actually invoke a broader, genome-wide program of differentiation-associated gene

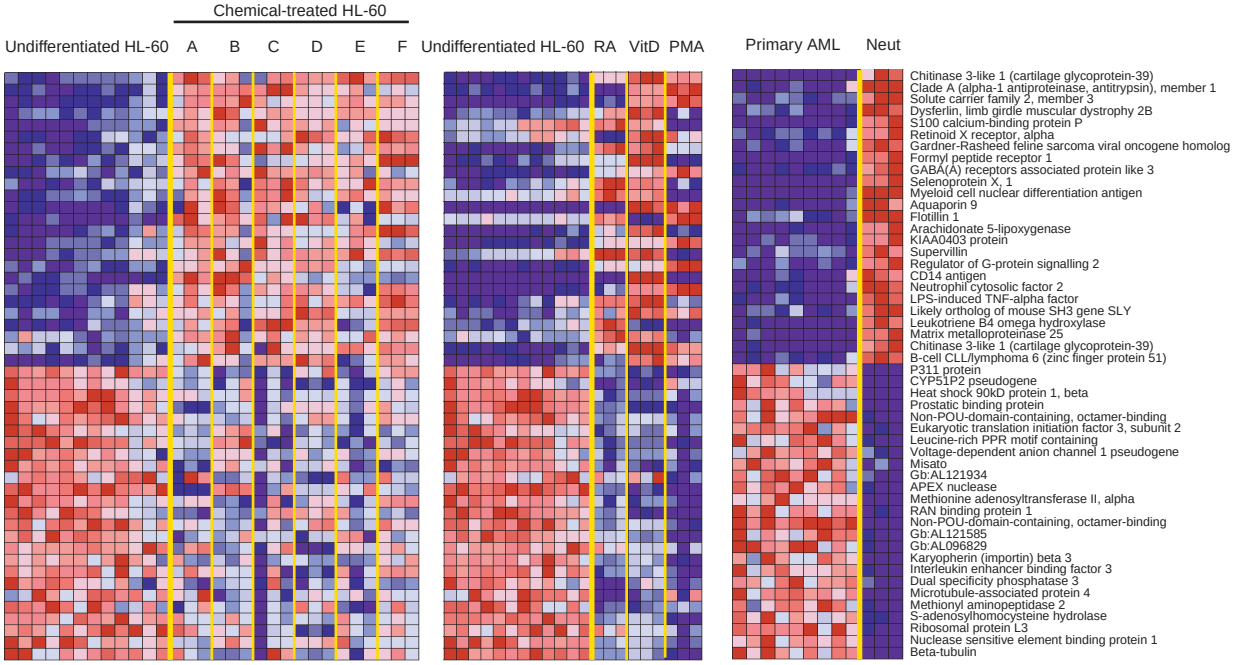

Figure 3 Whole-genome expression modulation. Gene expression consequences of candidate compounds. Gene expression profiling was done in triplicate, 5 d after treatment of HL-60 cells with compounds, and these patterns compared with the profiles distinguishing primary AML cells from normal donor neutrophils (Neut). Twelve HL-60 cell samples untreated or treated with 0.1% DMSO served as negative controls. Using the SNR metric (see **Supplementary Table 1i**,**1j** online), the genes distinguishing nine primary samples from individuals with AML from three normal neutrophil samples were identified and then reordered according to their degree of regulation in HL-60 cells by the candidate compounds. The top 25 genes in each direction are shown. Gene names or GenBank (Gb) accession numbers are shown at right. A, (R)-(-)-apomorphine HCl; B, 4,5-dianilinophthalimide; C, erythro-9-(2-hydroxy-3-nonyl) adenine HCl; D, 5-fluorouridine; E, 16-ketoestradiol; F, pergolide methanesulfonate. These genes are also shown in the space of undifferentiated HL-60 cells versus controls treated with ATRA (RA), 1,25-dihydroxyvitamin D3 (VitD) and PMA.

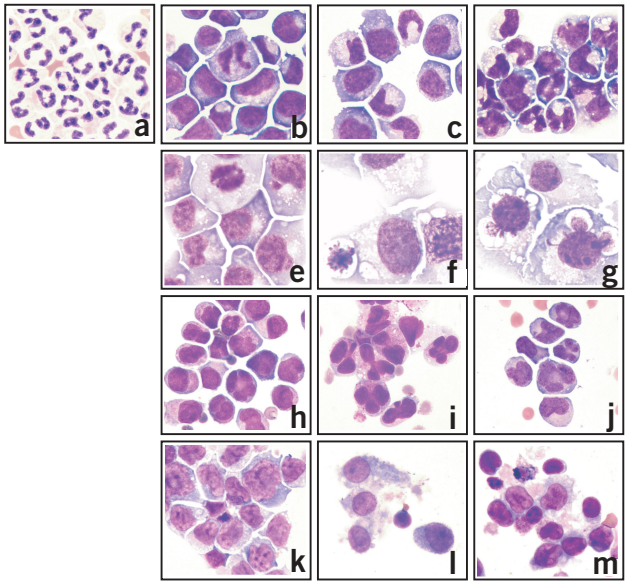

Figure 4 Morphological changes induced by chemical hits. May-Grunwald-Giemsa staining of (**a**) normal neutrophils, (**b**) HL-60 cells, (**c**) HL-60 cells treated with 70 μM 16-ketoestradiol for 5 d, (**d**) HL-60 cells treated with 1 μM ATRA for 5 d, (**e**) U937 cells, (**f**) U937 cells treated with 0.1 μM 5-fluorouridine for 5 d, (**g**) U937 cells treated with 70 μM erythro-9-(2-hydroxy-3-nonyl) adenine HCl for 5 d, (**h**) primary cells from an individual with APL, (**i**) primary cells from an individual with APL treated with 70 μM 16-ketoestradiol for 3 d, (**j**) primary cells from an individual with APL treated with 1 μm ATRA for 3 d, (**k**) primary cells from an individual with M1-AML, (**l**) primary cells from an individual with M1-AML treated with 70 μM 16-ketoestradiol for 5 d and (**m**) primary cells from an individual with M1-AML treated with 1 μM ATRA for 5 d.

expression. To answer this question, we treated HL-60 cells in triplicate with each of the eight compounds and obtained their gene expression profiles 5 d later using oligonucleotide microarrays containing probes for 22,283 human genes and expressed-sequence tags. We then compared the genes regulated by each of the compounds with the gene expression differences characterizing the original AML-versus-neutrophil and AML-versus-monocyte distinctions in primary cells. Six of the eight compounds induced a genome-wide change in gene expression that was statistically significantly similar to that seen in bona fide neutrophils, as determined by the Mantel test (**Fig. 3** and **Table 1**). The Mantel test is an unbiased, global measure of similarity, indicating that the six compounds induced a nonrandom program of gene expression consistent with differentiation. Notably, however, the magnitude of induction of the differentiation-associated gene expression pattern was weaker than that observed with the positive controls ATRA or PMA (**Supplementary Tables 1h–k** online).

To further confirm the biological consequence of compounds identified by GE-HTS, we first examined cells for the nuclear condensation and lobulation characteristic of polymorphonuclear neutrophils. For example, the estradiol derivative 16-ketoestradiol induced marked nuclear maturation in HL-60 cells (**Fig. 4a–d**). Notably, differentiating effects were not restricted to the HL-60 leukemia cell line. For example, erythro-9-(2-hydroxy-3-nonyl) adenine HCl and 5-fluorouridine induced macrophage morphology (**Fig. 4e–g**) in the myeloid leukemia cell line U937 (see also **Supplementary Table 1l** online). Furthermore, we observed compound effects in primary leukemia cells. For example, treatment with 16-ketoestradiol of primary leukemic blasts from an individual with microgranular variant acute promyelocytic leukemia induced neutrophilic morphology and whole-genome expression changes consistent with neutrophilic differentiation (**Fig. 4h–j**). Similarly, 16-ketoestradiol induced morphological evidence of macrophage maturation in leukemic blasts from an individual with type M1 AML, whereas ATRA had minimal effect (**Fig. 4k–m** and **Supplementary Tables 1m**, **n** online).

To extend these morphological observations, we next carried out three functional tests of differentiation. First, we showed that, as expected for terminally differentiating cells, many of the compounds (six of eight, 75%) inhibited cell proliferation at concentrations that induce the differentiation signature (**Table 1**). In addition, seven of eight candidate compounds scored positive in an NBT assay, consistent with their myeloid maturation (**Fig. 5a** and **Table 1**). Furthermore, we observed NBT reduction in primary cells from affected individuals. For example, 16-ketoestradiol induced NBT reduction in primary M1-AML cells, whereas ATRA did not ($P < 0.004$). We also found that five of eight candidate compounds induced engulfment of fluorescent carboxylate microspheres, consistent with phagocytic capacity of mature myeloid cells ($P < 0.03$; **Fig. 5b** and **Supplementary Note** online).

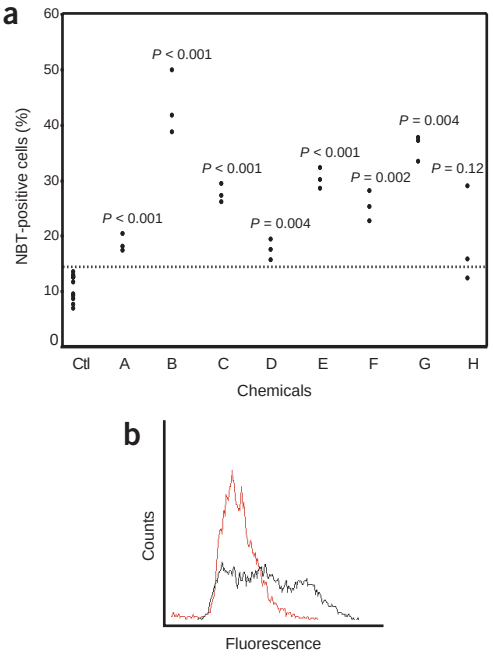

Figure 5 Functional changes induced by chemical hits. (**a**) NBT reduction assay. HL-60 cells were treated with compounds in triplicate for 6 d. Nine untreated HL-60 cell negative control samples were also evaluated. The percentage of blue cells was determined by light microscopy for at least 200 cells per sample. A, (R)-(-)-apomorphine HCl; B, 4,5-dianilinophthalimide; C, erythro-9-(2-hydroxy-3-nonyl) adenine HCl; D, 5-fluorouridine; E, 16-ketoestradiol; F, pergolide methanesulfonate; G, cyclazosin HCl; H, 1,10-phenanthroline. (**b**) Phagocytosis assay. HL-60 cells were treated with 4,5-dianilinophthalimide (25 μM) and compared with untreated HL-60 controls after 5 d. Fluorescent beads were then added for 1 h and cells were analyzed by fluorescence-activated cell sorting. The red spectra represents the untreated control cells, and the black the cells treated with 4,5-dianilinophthalimide.

These experiments indicate that all eight compounds that reproducibly triggered the differentiation signature induced at least one functional hallmark of bona fide differentiation, and that four of these were positive for all functional measures. Notably, phagocytosis and NBT reduction could be separated by some of the compounds. This suggests that chemical genomic screens, such as that described here, might be useful for the dissection of complex physiological processes. Our screen also yielded five false positives (0.3%), but each of these was quickly eliminated with repeat measurement of the signature. The GE-HTS method thus rapidly identified compounds that not only regulated the signature genes but also invoked the actual biological state transition of interest.

Although any chemical library can be used with GE-HTS, the one used here included many compounds with known mechanisms of action, thereby providing insight into the potential mechanism by which these compounds may induce differentiation. For example, five of the compounds identified in this screen (dimaprit dihydrochloride, pergolide methanesulfonate, (R)-(-)-apomorphine HCl, erythro-9-(2-hydroxy-3-nonyl) adenine HCl and the β-estradiol-related compound 16-ketoestradiol) increase intracellular cyclic AMP, which in turn potentiates myeloid differentiation[13–18]. The fact that the estrogen derivative 16-ketoestradiol had differentiation-promoting activity prompted us to examine structurally related compounds, including β-estradiol, 17-α-ethynylestradiol and 17-α-estradiol, all of which induced neutrophilic morphology and NBT reduction ($P < 0.02$; **Supplementary Table 1o** and **Supplementary Note** online), suggesting that it is the estrogenic component of these molecules that promotes differentiation, as previously suggested[19,20]. Two of the other compounds identified in this screen block pyrimidine synthesis through inhibition of either dihydrofolate reductase (aminopterin) or thymidylate synthetase (5-fluorouridine)[21,22]. Although these agents abrogate DNA synthesis and thus inhibited cell proliferation, other agents that similarly impede proliferation did not induce the differentiation signature, consistent with the idea that differentiation is not simply a default molecular program activated in the absence of proliferation[23]. Clearly, additional experiments are required to determine definitively the mechanism by which differentiation is induced by the compounds identified in this screen.

With the completion of the Human Genome Project, it is more straightforward to define cellular states based on genomic information. The challenge ahead is to bring chemistry to the systematic modulation of these states. The GE-HTS strategy described here is designed to provide a general approach to the discovery of state-modulating molecules identified solely by their ability to induce a surrogate gene expression signature. Such compounds could serve both as research reagents and as possible therapeutic leads. Although we have applied GE-HTS to leukemia, it could be applied, in principle, to any cellular response that affects gene expression. For example, the method could be used to identify compounds that abrogate physiologic processes (*e.g.*, cellular migration) or particular signaling pathways (*e.g.*, kinase cascades).

Unlike traditional assay techniques, such as those based on cellular phenotypes[1–3], reporter constructs[24] or antibody-based methods[25,26], GE-HTS does not require any specialized assays, reagents or assay customization. The gene expression signature definition, amplification and detection are entirely generic. Furthermore, unlike other gene expression-based methods designed to identify inhibitors of previously validated targets[27], such prior target validation is not required for GE-HTS; the gene expression signature simply serves as a surrogate for the biological state in question. In the leukemia differentiation case, we show that a modest number of genes are sufficient to capture a complex cellular response. Further investigation will, of course, be needed to identify the optimal type of gene signature. The use of more genes may increase specificity or enable the identification of compounds that trigger only part of a cellular response, and thereby help to dissect it.

An important goal for the decade ahead is the development of a wide range of chemical tools for use in modulating cellular states. The GE-HTS approach described here should prove useful for the systematic identification of compounds capable of modulating biological processes for which the key effectors are not yet known.

METHODS

Characterization of differentiation signatures. We obtained normal peripheral blood monocytes and neutrophils by Ficoll-Paque (Amersham Pharmacia Biotech) separation from three normal human donors provided by the Dana Farber Cancer Institute blood bank. We obtained primary samples from adults with AML from the Cancer and Leukemia Group leukemia bank as previously reported[28]. We isolated RNA using Trizol (GIBCO/BRL) following the manufacturer's instructions and prepared it for hybridization to Affymetrix HuFL microarrays containing 6,817 genes and expressed-sequence tags as previously reported[28].

After carrying out standard preprocessing steps, we selected differentiation signature genes using the signal-to-noise (SNR) statistic to rank the genes that correlated with the AML-versus-neutrophil and AML-versus-monocyte distinctions. SNR = $(\mu_{classo} - \mu_{class1})/(\sigma_{classo} + \sigma_{class1})$ where μ and σ represent the mean and the standard deviation of the expression, respectively, for each class. Permutation of the sample labels was done to compare these correlations with what would be expected by chance. We carried out 2,500 permutations to identify the differentiation signature genes that met statistical significance at the 80th percentile using GeneCluster2. We selected *GAPD* as the control gene based on the Affymetrix expression data and historical use of *GAPD* as a control. We confirmed the behaviors of the signature genes in the AML cell line HL-60 (American Type Culture Collection), which was differentiated in duplicate toward a neutrophil with 1 μM ATRA (Sigma) for 0, 24, 72 and 120 h or toward a monocyte-macrophage-like cell with 10 nM PMA (Sigma) for 0, 4, 12 and 24 h. We extracted, labeled and hybridized RNA to Affymetrix HuFL microarrays. Full details of target preparation and microarray analyses are described in **Supplementary Note** online.

High-throughput RT-PCR amplification of signature genes. We grew HL-60 cells in 384-well culture plates in 40 μl of medium (RPMI 1640 (Cellgro) with 10% fetal bovine serum (Sigma) and 1% penicillin-streptomycin (Cellgro)) at 0.45×10^6 cells ml^{-1}. We lysed cells with 45 μl per well of a mixture containing a hypotonic detergent, dithiothreitol and RNAse inhibitor. We then transferred 15 μl of lysate and 6 μl of a 2.5× binding buffer to a 384-well oligo-dT-coated plate. After washing, we carried out reverse transcription in a 20-μl M-MuLV reaction at 37 °C for 1.5 h. We purchased lysis buffers, 384-well custom-coated oligo-dT plates, wash buffers and M-MuLV from Pierce and used them according to a modified version of their Express Direct mRNA Capture and RT-PCR system.

After carrying out multiplexed PCR, we treated 5 μl of the reaction with shrimp alkaline phosphatase (Sequenom) to inactivate remaining dNTPs. We designed SBE probes of 16–21 nucleotides with an annealing temperature of 50–55 °C. We carried out five-plex SBE reactions in 9-μl reaction volumes with 1× Thermosequenase buffer, 2.7 μM of each primer, 0.2 mM of each ddNTP (Sequenom) and 0.58 units per reaction of Thermosequenase (Sequenom) in an MJ 384-well Thermocycler (92 °C for 2 min, 40 cycles of 92 °C for 20 s, 50 °C for 30 s). We then treated the SBE product with a resin to remove residual salt and loaded it onto a 3-hydroxypicolinic acid–based matrix pad (SpectroCHIP, Sequenom) with a Spectropoint robot (RoboDesign). SpectroCHIPS were analyzed with a Bruker Biflex MALDI-TOF mass spectrometer (SpectroREADER), and spectra were processed using SpectroTYPER software (Sequenom). For each fragment, the peak intensity corresponding to the expected mass of the extended primer was recorded and its value adjusted for background noise. (See **Supplementary Note** online for full details of methods and primer sequences.)

Screening methods. We grew HL-60 cells at 0.45×10^6 cells ml^{-1} in 40-μl volumes in 384-well culture plates. We treated cells at a uniform concentration with 4 μg ml^{-1} (corresponding to ~10 μM) of each compound dissolved in 0.1% dimethylsulfoxide (DMSO). Details of the small molecule library can be found in **Supplementary Note** online. After 3 d, we extracted RNA and carried out RT-PCR in 20-μl reaction volumes as described above with five primer pairs: *GAPD*, *IL1RN*, *SPP1*, *NCF1* and *ORM1*. We then carried out SBE and MALDI-TOF mass spectrometry as described above and calculated the Neutrophil and Monocyte Scores as detailed in **Supplementary Note** online.

Confirmation of candidate compounds: signature confirmation. We treated HL-60 cells in triplicate for 5 d with candidate compounds at the following concentrations: (R)-(-)-apomorphine HCl, 5 μM; 8-(3-chlorostyryl) caffeine, 75 μM; cyclazosin HCl, 7.5 μM; 4,5-dianilinophthalimide, 30 μM; erythro-9-(2-hydroxy-3-nonyl) adenine HCl, 70 μM; 5-fluorouracil, 1 μM; 5-fluorouridine, 0.1 μM; 16-ketoestradiol, 70 μM; α-methyl-L-p-tyrosine, 100 μM; pergolide methanesulfonate, 50 μM; 1,10- phenanthroline, 0.8 μM; (-) scopolamine methyl bromide, 75 μM; and sulmazole, 70 μM. We purchased all compounds from Sigma, except cyclazosin HCl, which was provided by D. Giardina (Department of Chemical Sciences, University of Camerino, Camerino, Italy). We determined effective concentrations by preliminary evaluation of compound-induced differentiation and growth inhibition. We used untreated HL-60 cells and HL-60 cells treated with 0.1% DMSO vehicle as negative controls. We isolated RNA from cells treated with candidate compounds and hybridized it to Affymetrix U133A microarrays (**Supplementary Note** online). We compared the mean expression value for each marker gene in the compound-treated cells with that of the negative controls by evaluation of relative induction and a one-tailed *t*-test analysis of statistical significance.

Whole-genome evaluation of compounds. We used the Mantel test, a non-parametric randomization-based procedure that estimates the correlation between two distance matrices, to assess whether the chemicals induced whole-genome changes consistent with differentiation[29]. For a given gene expression data set X and corresponding class labels (*e.g.*, compound-treated versus untreated HL-60 cells or AML blasts versus neutrophils), we calculated the distance of each feature from the class labels using the SNR statistic as follows:

$$X_i = \frac{\mu_{i1} - \mu_{i2}}{\sigma_{i1} + \sigma_{i2}}$$

where μ_{i1} represents the mean expression of samples from class 1 for feature i and σ_{i1} represents the standard deviation of class 1 for feature i. Similarly, we calculated the SNR statistic for the second set of samples Y. The elements of vector X and vector Y correspond to the same set of objects (genes). We computed the Pearson correlation between the corresponding elements of the two vectors, producing the Mantel correlation R_m as follows:

$$R_m = \frac{\sum_{i=1}^{N} X_i Y_i - \frac{\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{N}}{\sqrt{\left(\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right)\left(\sum_{i=1}^{N} Y_i^2 - \frac{\left(\sum_{i=1}^{N} Y_i\right)^2}{N}\right)}}$$

We used R_m as the reference value in the Mantel test. We computed significance levels by randomly permuting the elements of one of the vectors to produce a permuted vector X^* and, as before, computed $R_m{}^*$ between X^* and Y. We repeated the permutation-computation steps 2,500 times and used the resulting distribution to estimate the P value by determining the proportion of $R_m{}^*$ values that exceeded R_m (**Supplementary Note** online).

NBT reduction assay. We treated cells with compounds for 6 d and then incubated them at 37 °C for 1 h in medium containing 10% fetal bovine serum, 0.1% NBT (Sigma) and 1 μg ml^{-1} PMA (Sigma). We determined the percentage of blue cells by light microscopy for at least 200 cells per sample and used a one-tailed *t*-test analysis to compare this percentage with that for untreated cells[8]. The percentage of NBT-positive cells in untreated cells versus cells treated with 0.1% DMSO was not statistically significantly different (**Supplementary Note** online).

Phagocytosis assay. After 5 d of treatment with candidate compounds, we incubated HL-60 cells for 1 h with 0.026% fluorescent latex beads (Fluoresbrite Carboxylate 0.75 micron microspheres, Polysciences) and then washed them three times with phosphate-buffered saline. We analyzed fluorescent uptake by fluorescence-activated cell sorting with a Becton Dickinson FACScan and CELLQuest analytical software[30]. We used laser excitation of 488 nm and detected fluorescent emission of 530/30 nm band pass. Cells without beads established the gate for live cells using forward and side scatter patterns, and we then used untreated cells incubated with beads to establish the background fluorescence. We set an M1 gate at 5% for the untreated and vehicle-treated control cells. An average of the five replicates was calculated. We compared compound-treated cells with controls using a one-tailed *t*-test assuming two groups with unequal variance (**Supplementary Note** online).

Blast analysis of primary cells from individuals with AML. We obtained samples from individuals with AML from the Children's Hospital of Boston and Dana Farber Cancer Institute with approval of the Dana Farber/Harvard Cancer Center Internal Review Board and informed consent from the affected individuals or a parent at time of diagnosis. We obtained one sample from a leukopharesis of an individual with APL, t(15;17) and a white-blood-cell count of 185,000 with 97% myeloblasts at diagnosis. The other sample was from the peripheral blood of an individual with M1-AML with monocytic features, a white-blood-cell count of 37,200 with 74% percent myeloblasts and trisomy 8. We treated leukemic cells in duplicate with 1 μM ATRA or 70 μM 16-ketoestradiol and evaluated them with May-Grunwald-Giemsa staining daily. After 5 d, we extracted RNA and hybridized it to U133A Affymetrix microarrays (**Supplementary Note** online).

URLs. GeneCluster2 is available at http://www.broad.mit.edu/cancer/software/software.html. Raw microarray data are available at http://www.broad.mit.edu/cancer/pub/GE-HTS_leuk or http://www.ncbi.nlm.nih.gov/geo/ and described in **Supplementary Table 1p** online.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank members of the laboratory of T.R.G., the Broad Institute Cancer Program and, in particular, B. Blumenstiel, R. Byers, M. Defelice, S. Dolma, K. El-Hayak, M. Fleming, S. Gabriel, S. Lessnick, D. Peck, M. Porcionatto, A. Subramanian and P. Tamayo for their contributions and E. Lander for critical reading of the manuscript. This work was supported in part by the National Institutes of Health, the Leukemia and Lymphoma Society of America and the Howard Hughes Medical Institute. B.R.S. was funded in part by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and by the National Cancer Institute.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 18 December 2003; accepted 21 January 2004
Published online at http://www.nature.com/naturegenetics/

1. Peterson, R.T., Link, B.A., Dowling, J.E. & Schreiber, S.L. Small molecule developmental screens reveal the logic and timing of vertebrate development. *Proc. Natl. Acad. Sci. USA* **97**, 12965–12969 (2000).
2. Mayer, T.U. *et al.* Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971–974 (1999).
3. Grozinger, C.M., Chao, E.D., Blackwell, H.E., Moazed, D. & Schreiber, S.L. Identification of a class of small molecule inhibitors of the sirtuin family of NAD-dependent deacetylases by phenotypic screening. *J. Biol. Chem.* **276**, 38837–38843 (2001).
4. Clemons, P.A. *et al.* A one-bead, one-stock solution approach to chemical genetics: part 2. *Chem. Biol.* **8**, 1183–1195 (2001).
5. Kakizuka, A. et al. Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML. *Cell* **66**, 663–674 (1991).
6. Huang, M.E. *et al.* Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood* **72**, 567–572 (1988).
7. Seiter, K. *et al.* Clinical and laboratory evaluation of all-trans retinoic acid modulation of chemotherapy in patients with acute myelogenous leukaemia. *Br. J. Haematol.*

108, 40–47 (2000).
8. Sokoloski, J.A., Blair, O.C. & Sartorelli, A.C. Alterations in glycoprotein synthesis and guanosine triphosphate levels associated with the differentiation of HL-60 leukemia cells produced by inhibitors of inosine 5′-phosphate dehydrogenase. *Cancer Res.* **46**, 2314–2319 (1986).
9. Yeakley, J.M. *et al.* Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* **20**, 353–358 (2002).
10. Yang, L., Tran, D.K. & Wang, X. BADGE, Beads Array for the Detection of Gene Expression, a high-throughput diagnostic bioassay. *Genome Res.* **11**, 1888–1898 (2001).
11. Root, D.E., Flaherty, S.P., Kelley, B.P. & Stockwell, B.R. Biological mechanism profiling using an annotated compound library. *Chem. Biol.* **10**, 881–892 (2003).
12. Bodner, A.J., Ting, R.C. & Gallo, R.C. Induction of differentiation of human promyelocytic leukemia cells (HL-60) by nucleosides and methotrexate. *J. Natl. Cancer Inst.* **67**, 1025–1030 (1981).
13. Kalinyak, K.A., Sawutz, D.G., Lampkin, B.C., Johnson, C.L. & Whitsett, J.A. Effects of dimaprit on growth and differentiation of human promyelocytic cell line, HL-60. *Life Sci.* **36**, 1909–1916 (1985).
14. Fici, G.J., Wu, H., VonVoigtlander, P.F. & Sethy, V.H. D1 dopamine receptor activity of anti-parkinsonian drugs. *Life Sci.* **60**, 1597–1603 (1997).
15. Perachon, S., Schwartz, J.C. & Sokoloff, P. Functional potencies of new antiparkinsonian drugs at recombinant human dopamine D1, D2 and D3 receptors. *Eur. J. Pharmacol.* **366**, 293–300 (1999).
16. Duncan, G.S., Wolberg, G., Schmitges, C.J., Deeprose, R.D. & Zimmerman, T.P. Inhibition of lymphocyte-mediated cytolysis and cyclic AMP phosphodiesterase by erythro-9-(2-hydroxy-3-nonyl)adenine. *J. Immunopharmacol.* **4**, 79–100 (1982).
17. Dubey, R.K. *et al.* Estradiol inhibits smooth muscle cell growth in part by activating the cAMP-adenosine pathway. *Hypertension* **35**, 262–266 (2000).
18. Guillemin, M.C. *et al.* In vivo activation of cAMP signaling induces growth arrest and differentiation in acute promyelocytic leukemia. *J. Exp. Med.* **196**, 1373–1380 (2002).
19. Mountford, J.C. *et al.* Estrone potentiates myeloid cell differentiation: a role for 17 beta-hydroxysteroid dehydrogenase in modulating hemopoiesis. *Exp. Hematol.* **27**, 451–460 (1999).
20. Shim, G.J. *et al.* Disruption of the estrogen receptor beta gene in mice causes myeloproliferative disease resembling chronic myeloid leukemia with lymphoid blast crisis. *Proc. Natl. Acad. Sci. USA* **100**, 6694–6699 (2003).
21. McBurney, M.W. & Whitmore, G.F. Mechanism of growth inhibition by methotrexate. *Cancer Res.* **35**, 586–590 (1975).
22. Reyes, P. & Heidelberger, C. Fluorinated pyrimidines. XXVI. Mammalian thymidylate synthetase: its mechanism of action and inhibition by fluorinated nucleotides. *Mol. Pharmacol.* **1**, 14–30 (1965).
23. Ben-Baruch, N. *et al.* c-myc Down-regulation in suramin-treated HL60 cells precedes growth inhibition but does not trigger differentiation. *Mol. Pharmacol.* **46**, 73–78 (1994).
24. Baker, K. *et al.* Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **99**, 16537–16542 (2002).
25. Stockwell, B.R., Haggarty, S.J. & Schreiber, S.L. High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. *Chem. Biol.* **6**, 71–83 (1999).
26. Barrie, S.E. *et al.* High-throughput screening for the identification of small-molecule inhibitors of retinoblastoma protein phosphorylation in cells. *Anal. Biochem.* **320**, 66–74 (2003).
27. Johnson, P.H. *et al.* Multiplex gene expression analysis for high-throughput drug discovery: screening and analysis of compounds affecting genes overexpressed in cancer cells. *Mol. Cancer Ther.* **1**, 1293–1304 (2002).
28. Golub, T.R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
29. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
30. Harvath, L. & Terle, D.A. Assay for phagocytosis. *Methods Mol. Biol.* **115**, 281–290 (1999).

Corrigendum: Gene expression–based high-throughput screening (GE-HTS) and application to leukemia differentiation

K Stegmaier, K N Ross, S A Colavito, S O'Malley, B R Stockwell & T R Golub
Nat. Genet. **36, 257–263 (2004).**

The GEO accession number for the array data is GSE995.

Erratum: Assessing the validity of the association between the SUMO4 M55V variant and risk of type 1 diabetes

D J Smyth, J M M Howson, C E Lowe, N M Walker, A C Lam, S Nutland, J Hutchings, E Tuomilehto-Wolf, J Tuomilehto, C Guja, C Ionescu-Tîrgovişte, D E Undlien, K S Rønningen, D Savage, D B Dunger, R C J Twells, W L McArdle, D P Strachan & J A Todd
Nat. Genet. **37, 110–111 (2005).**

The authors' affiliations were mistakenly omitted. A complete list of authors and their affiliations follows.

Deborah J Smyth[1], Joanna M M Howson[1], Christopher E Lowe[1], Neil M Walker[1], Alex C Lam[1], Sarah Nutland[1], Jayne Hutchings[1], Eva Tuomilehto-Wolf[2], Jaakko Tuomilehto[2–4], Cristian Guja[5], Constantin Ionescu-Tîrgovişte[5], Dag E Undlien[6], Kjersti S Rønningen[7], David Savage[8], David B Dunger[9], Rebecca C J Twells[1], Wendy L McArdle[10], David P Strachan[11] & John A Todd[1]

[1]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Rd, Cambridge, UK. [2]Diabetes and Genetic Epidemiology Unit, National Public Health Institute, University of Helsinki, Helsinki, Finland. [3]Department of Public Health, University of Helsinki, Helsinki, Finland. [4]South Ostrobotnia Central Hospital, Seinäjoki, Finland. [5]Clinic of Diabetes, Institute of Diabetes, Nutrition and Metabolic Diseases 'N. Paulescu', Bucharest, Romania. [6]Institute and Department of Medical Genetics, Ulleval University Hospital, University of Oslo, Oslo, Norway. [7]Laboratory of Molecular Epidemiology, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway. [8]Department of Medical Genetics, Queen's University Belfast, Belfast City Hospital, Belfast, Northern Ireland. [9]Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK. [10]Avon Longitudinal Study of Parents and Children, University of Bristol, Bristol, UK. [11]Department of Community Health Sciences, St George's Hospital Medical School, London, UK. Correspondence should be addressed to J.A.T. (john.todd@cimr.cam.ac.uk).

Corrigendum: Low LDL cholesterol in African Americans resulting from frequent nonsense mutations in *PCSK9*

J Cohen, A Pertsemlidis, I K Kotowski, R Graham, C K Garcia & H H Hobbs
Nat. Genet. **37, 161–165 (2005).**

The 25-year-old offspring of the proband in Figure 3a is a woman.